LU TP 98-1
February 22, 1998

Clustering ECG Complexes Using Hermite Functions and Self-Organizing Maps

Martin Lagerholm[1], Carsten Peterson[2]

Complex Systems Group, Department of Theoretical Physics
Lund University, Sölvegatan 14A, S-223 62 Lund, Sweden

Guido Braccini, Lars Edenbrandt[3]

Department of Clinical Physiology, Lund University, S-221 85 Lund, Sweden

Leif Sörnmo[4]

Signal Processing Group, Department of Applied Electronics, Lund University,
Box 118, S-221 00 Lund, Sweden

Submitted to *IEEE Transactions on Biomedical Engineering*

Abstract:

An integrated method for identifying and classifying ECG complexes in the MIT-BIH Arrhythmia Database is presented. The QRS complexes are extracted and the RR intervals are calculated in an autonomous way. Each complex is subsequently decomposed into Hermite basis functions, from which the resulting coefficients and widths are used to represent the beats. By means of this representation, unsupervised self-organizing neural networks are employed to cluster the data into 25 groups. The resulting clusters exhibit a very low degree of misclassification (1.5%). The approach outperforms published supervised learning methods using the same data. The method is also successfully compared with a more conventional template cross-correlation clustering method.

[1]martin.lagerholm@thep.lu.se
[2]carsten.peterson@thep.lu.se
[3]lars.edenbrandt@klinfys.lu.se
[4]leif.sornmo@tde.lth.se

1 Introduction

Computer-based methods for analysis and interpretation of ECGs have been subject to intense research for nearly four decades. Electrocardiographs that perform sophisticated signal processing and diagnostic interpretation of signals are today widely spread. Artificial neural networks (**ANN**) is one of the most recent techniques in this field and have primarily been considered for classification of ECGs into different diagnostic groups [1] [2]. It has been shown that ANNs for specific issues can perform better than both experienced cardiologists and ruled-based criteria, e.g. in detecting acute myocardial infarction from the ECG [3]. A first generation of ANNs have also been implemented in commercial electrocardiographs.

Visual analysis of long-term (24h) ECG is tedious and operator dependent. Computer techniques have been developed in order to facilitate the visual analysis, e.g. by condensed print-outs of various signals and trends. With this type of presentation the operator usually can analyze a 24-hour recording in 20-40 minutes provided that no complex arrhythmias exist. It is obvious that automated systems for detection of arrhythmias considerably reduces the amount of time the operator needs to spend. Several commercial systems are available for long-term ECG analysis. However, their performances deteriorate markedly when noise and artefacts are present and, as a consequence, an excessive number of beat classes is created which require considerable manual editing.

Beat classification based on ANN requires that a set of descriptive measurements first has been extracted from the ECG. Several approaches have been taken to this problem of which perhaps the straightforward use of time samples from the ECG signal is most frequently used [4] [5] [6]. Another approach is to use a set of heuristic features e.g. measurements of QRS wave amplitudes and durations [7]. While both these approaches are likely to perform well in recordings acquired during favourable conditions, the classification performance will severely degrade in noisy recordings. A more robust approach to the feature extraction problem is to use a basis function representation of the ECG. This technique was suggested for ECG analysis already in the 60's when Laguerre orthogonal functions were proposed [8]. Later, the Karhunen-Loeve (K-L) expansion, which provides an optimal signal representation in the mean square error sense, was found to be suitable for this purpose, e.g. [9] [10]. It is well-known that the K-L basis functions constitute an orthonormal set and therefore each coefficient in the expansion represents independent information. Based on the shape of the most significant K-L basis functions, the orthogonal Hermite functions were later proposed as a useful parametric model of the QRS complex [11]. An advantage with the Hermite model is that it includes a width parameter. Recently, this model was studied within the context of beat classification [12], and data compression [13].

In this paper, we develop a method for unsupervised characterization of ECG signals. The aim is to partition the beats into clusters that represent central features of the data such that similarity structures between clusters are preserved. Our approach involves Hermite function representation of the QRS complexes in combination with self-organized clustering. In contrast to feedforward supervised ANN learning algorithms, which are trained to identify predefined features in the data, self-organizing networks (**SOM**) organize themselves according to the natural structure of the data – no external teacher is called for [14]. Here, we use a variant of an SOM where neighbours are updated according to a soft-max (Potts) prescription. The input data vector consists of the Hermite coefficients, the width parameter and the lengths of adjacent RR intervals.

Recently, an ANN-based method for beat classification was presented which embraces the "mixture-of-expert" (**MOE**) principle [15]. The present work differs distinctly in several respects to that work, e.g. whereas our approach handles all steps from QRS detection to beat classification, the work in [15] is focused on patient-adaptable techniques that require initial labeling of beats. To the extent it is feasible, our results are compared to those in [15] using the MIT-BIH Arrhythmia database [16].

This paper is organized as follows. In Section 2 we briefly describe the MIT-BIH Arrhythmia database. Our preprocessing techniques, QRS detection and Hermite functions decomposition, are developed in Section 3. Section 4 describes the essentials of the self-organizing map and Section 5 presents the results. The correlation-based method is described in Section 6. Finally, conclusions and outlook can be found in Section 7.

2 ECG database

A total of 48 ECG recordings were studied from the MIT-BIH arrhythmia database [16]. Each recording has a duration of 30 minutes and includes two leads, denoted A and B respectively, which are chosen among the modified limb lead II and the modified leads V1, V2, V4, or V5. All beats in the database were annotated by two cardiologists: approximately 70% of the beats were classified as normals while the remaining types of beats were divided into 16 different kinds of abnormal beats. Since we focus on the QRS part of the signal, the "p" beats only containing the P wave are omitted. Hence we only use 15 abnormal beat types (see Table 1). When comparing our procedure to the annotations in Table 1, we utilize all the 15 different categories rather than lumping them together as normal/abnormal beats.

The ECG signals were resampled from the original 360 Hz sampling rate to 1 kHz in order to be compatible with the software used for high-resolution ECG analysis, including an algorithm for QRS detection [17].

3 Preprocessing

The coefficients of the Hermite basis function representation are determined from a highpass filtered ECG signal in order to avoid poor fits due to the presence of very low-frequency components (linear-phase filtering with cut-off frequency at 0.3 Hz). For each QRS complex, the fiducial point is related to the peak location in a signal obtained as the sum of envelopes from the two leads [17]. Next, the ECG signal is extracted from a t_0=200 ms window centered around the fiducial point. The window length is selected on the basis of an exploratory investigation with respect to Hermite function decomposition efficiency and classification performance. The overall efficiency of this extraction procedure when comparing with the database annotation file is high – 99.7%. The 0.3% discrepancy is broken down into beat types in Table 1.

Each QRS complex is represented by a $2t_0$-dimensional vector $\vec{x} = \{x_t\}$ which is expanded onto a linear combination of Hermite basis functions $\phi_t^{(n)}(\sigma)$, n=0,1,...,N-1, plus an unknown error term

code	beat type	Annotation	Our Method	Δ (%)
N	Normal beat	75053	74854	0.3
L	Left bundle branch block	8074	8060	0.2
R	Right bundle branch block	7259	7239	0.3
A	Atrial premature beat	2544	2538	0.2
a	Abberated atrial premature beat	150	109	27.0
J	Nodal (junctional) premature beat	83	83	0.0
S	Supraventricular premature beat	2	2	0.0
V	Ventriculare premature beat	7129	7110	0.3
F	Fusion of ventricular and normal beat	803	798	0.6
b	Ventricular flutter wave	472	458	0.3
e	Atrial escape beat	16	16	0.0
j	Nodal (junctional) escape beat	229	228	0.4
E	Ventricular escape beat	106	106	0.0
P	Paced beat	7028	7014	0.2
f	Fusion of paced and normal beat	982	979	0.3
Q	Unclassifiable beat	33	33	0.0

Table 1: The 16 different types of beats from the MIT-BIH Arrhythmia Database with annotations. Also shown is how these are identified with our ECG signal extraction method and the differences (Δ) with the database annotation. The poor detection rate of the a-beats was due to the fact that several very low amplitude PVCs in record 201 where missed by the QRS detector.

$e_t(\sigma)$ [11] [12],

$$x_t = \sum_{n=0}^{N-1} c_n(\sigma)\phi_t^{(n)}(\sigma) + e_t(\sigma) \tag{1}$$

where

$$\phi_t^{(n)}(\sigma) = \frac{1}{\sqrt{\sigma 2^n n! \sqrt{\pi}}} e^{-t^2/2\sigma^2} H_n(t/\sigma) \tag{2}$$

and $H_n(t/\sigma)$ are the Hermite polynomials. The width σ approximates the half-power duration. With $H_0(x) = 1$ and $H_1(x) = 2x$, the Hermite polynomials are recursively given by

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x) \tag{3}$$

Hence each complex is represented by N+1 parameters; σ and $c_n(\sigma)$ (n=0,1,...,N-1). The first six Hermite functions are shown in Fig. 1. The Hermite functions form an orthonormal basis

$$\sum_{t=-\infty}^{\infty} \phi_t^{(n)}(\sigma)\phi_t^{(m)}(\sigma) = \delta_{mn} \tag{4}$$

for all values of σ, which allows for a cost-effective calculation of $c_n(\sigma)$ in x_t in (1).

For a finite window size $\phi_n(\sigma)$ does not obey (4). However, if $\phi_n(\sigma)$ is close to zero outside the window it is still a good approximation. In particular, this problem occurs for large n and σ, (see Fig. 1); for a given window size, N imposes an upper limit σ_0 on σ. On the other hand, for a good representation of x_t, σ_0 should not be too small. Fortunately, fairly small values of N, i.e. 3 or 4, allow for good beat representation. A window size of $2t_0 = 400$ ms, including a 100 ms zero padding

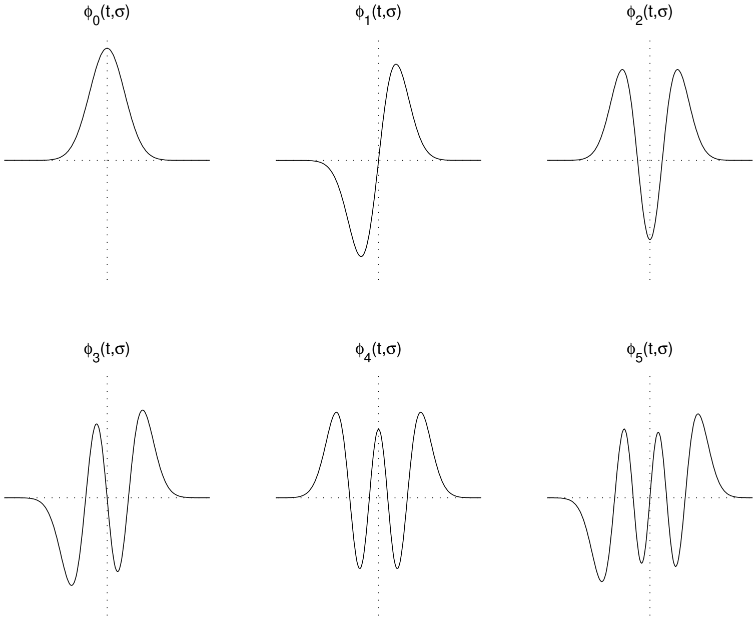

Figure 1: The first six Hermite functions (2,3) as functions of t.

on each side of the QRS complex, is in this respect a suitable choice. Defining $\phi_n(\sigma)$ as being close enough to zero outside the window if

$$|\phi^n_{-t_0}(\sigma)| = |\phi^n_{t_0}(\sigma)| < \frac{1}{10} \max_{t\in[-t_0,t_0]} |\phi^{(n)}_t(\sigma)| \tag{5}$$

one obtains for $N = 3$, 4, 5 and 6 the maximum widths $\sigma_0 = 62$, 55, 51 and 47 ms respectively. These numbers are sufficiently large for good representation of the beats.

For a given σ, the coefficients $c_n(\sigma)$ are determined by minimizing the summed square error

$$\sum_t |e_t(\sigma)|^2 = \sum_t |x_t - \sum_n c_n(\sigma)\phi^{(n)}_t(\sigma)|^2. \tag{6}$$

The minimum of (6), for a fix σ, is conveniently evaluated by means of the orthogonality property (4), given by

$$c_n(\sigma) = \vec{x} \cdot \vec{\phi}^{(n)}(\sigma) \tag{7}$$

The optimal combination of σ and the coefficients is determined by stepwise incrementation of σ up to σ_0 and then recomputation of (7) rather than fitting all parameters simultaneously in order to have a fast decomposition. The use of (7) significantly reduces the computational demand when decomposing the signals in Hermite functions. This represents an asset in real-time clustering.

As a relative error measure we use

$$\epsilon = \frac{\vec{e}(\sigma)^2}{\vec{x}^2} \tag{8}$$

Hermite functions (N)	3	4	5	6
$\epsilon(\%)$	9.7	6.8	5.5	4.5
$\langle\sigma\rangle(\%)$	19.1	19.0	18.7	18.4

Table 2: The relative error ϵ and $\langle\sigma\rangle$ averaged over both leads for N=3, 4, 5 and 6 Hermite functions respectively.

	N	L	R	A	a	J	S	V	F	b	e	j	E	P	f	Q
ϵ (%)	5	2	8	9	7	5	4	5	3	6	3	4	4	14	11	19

Table 3: The relative error ϵ for the different beat types when decomposing lead A and B into 5 Hermite basis functions.

In general, beats are well represented even with a small number of Hermite functions. This is confirmed both by visual inspection and by the fact that ϵ is low (see Table 2). The P and f beat types, which are pacemaker related, and Q (unclassifiable) are less suitable for representing with a few Hermite functions. However, the variance of ϵ between different types of beats is clearly acceptable when using five Hermite functions are used (see Table 3). Figure 2 exemplifies the representation of a normal beat with 3, 4, 5 and 6 Hermite functions. As will be shown later (Table 5) the performance of the SOM using 4, 5, and 6 Hermite functions representation are approximately the same. We settle for 5 in what follows.

Apart from the above parameters related to the Hermite function representation, local rhythm

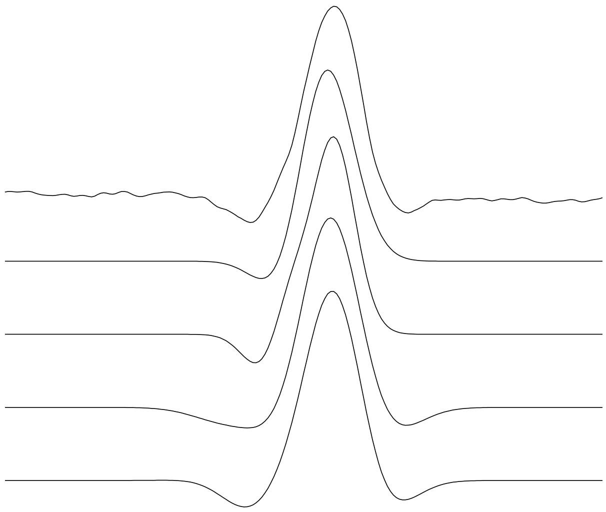

Figure 2: An example of Hermite function representation. The upper curve is the original beat (lead A in beat #1 of record 100). The subsequent four curves represent decomposition using 3, 4, 5 and 6 Hermite functions respectively.

information will also be used as input to the SOM. The following two features are considered:

$$\begin{aligned} R^- &\equiv R_i^- = R_i^* - R_{i-1}^* \\ R^+ &= \Psi[(R_{i+1}^- - R_i^-) - (R_i^- - R_{i-1}^-)] \end{aligned} \tag{9}$$

where R_i^* is the occurrence time for the i:the beat, $\Psi(x)$=x if $x > 0$, otherwise $\Psi(x)$=0. In this way R^+ does not depend on the heart rate, because it is significantly different from zero only if the i:th beat is premature. This is an important mark for atrial beats.

4 Self-Organizing Map

4.1 Self-Organizing Networks

Whereas feedforward supervised ANN learning algorithms are aimed at identifying predefined features in the data, SOMs operate in a different way. The desired feature values are not known beforehand and the network must organize itself according to the natural structure of the data. Hence this approach falls within the family of clustering methods.

Topological Maps

The idea is to map a set of K-dimensional input vectors $\vec{\xi}$ onto a two-dimensional array of output nodes h_j (j=1,...,M), such that the latter capture key features of the input vectors. Each output node correspond to a position in input space given by $\vec{\omega}_j$. The distance between h_j and input vector $\vec{\xi}$ is defined as

$$d_j = \sum_k^K (\omega_{jk} - \xi_k)^2. \tag{10}$$

where the index labeling the different input vectors is suppressed. In cases, where each input vector is normalized, this simple measure is sufficient. For each input vector $\vec{\xi}$, a "winner" h_{i^*} is chosen as the one closest to $\vec{\xi}$, i.e.

$$d_{i^*} = \min_j d_j \tag{11}$$

Input vectors having h_{i^*} as a winner are said to "belong" to cluster (output node) i^*. In what follows a cluster will also be referred to as a *box*. In the simplest form of self-organization, "winner-takes-all", the weights ω_{mk} are updated such that h_m will become more sensitive to the presented pattern, similar to standard clustering algorithms. This is accomplished by

$$\begin{aligned} \Delta\omega_{mk} &= \eta(\xi_k - \omega_{mk}) \quad &\text{if} \quad m = i^* \\ \Delta\omega_{mk} &= 0 \quad &\text{if} \quad m \neq i^* \end{aligned} \tag{12}$$

where η is the step size towards the input vector.

It is often desirable to have a topological order among the output units h_j, such that neighbouring units in the output space will respond to adjacent points in the input space. A commonly used technique to achieve such a *topological map* is by updating units close to the winner node in the same way as the winner node [14] but with a step that in addition to η also depends upon the

	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	ξ_8	ξ_9	ξ_{10}	ξ_{11}	ξ_{12}
lead A	c_1	c_2	c_3	c_4	σ							
lead B						c_1	c_2	c_3	c_4	σ		
lead A/B											R^-	R^+

Table 4: ξ-representation.

distance to the winner, r (in the output space). In this way neighbouring units will end up with similar weight vectors. We do this using *Potts* neurons, where the winner "shares" its step with its neighbours. The Potts neuron, $\vec{v}_{i^*j}$, encodes the relative strengths according to

$$v_{i^*m} = \frac{e^{-r^2_{i^*m}/T}}{\sum_k e^{r^2_{i^*k}/T}}. \tag{13}$$

The distance r_{ij} (in the output space) between the output units defines the topology of the network. The artificial "temperature" or width T is decreased (annealed) as learning proceeds. The update rule (12) is then modified to

$$\Delta\omega_{mk} = \eta(\xi_k - \omega_{mk})v_{i^*m} \;\text{ for all }\; m \tag{14}$$

Note that

$$\sum_j v_{ij} = 1, \tag{15}$$

expressing the "share" philosophy. For $T \rightarrow 0$ (14) reduces to the "winner-take-all" update rule (12). Suitable parameter choices for our procedure are found below.

4.2 The QRS Map

4.2.1 Input Structure

We map our preprocessed QRS data, denoted by ξ_k, onto 5×5 maps using (10) and (11). The ξ_k-representation is shown in Table 4 where c_i and σ refer to the Hermite decomposition for each lead. Since the components of the input vector (c_n, σ, R^- and R^+) have different origins one does not *a priori* know their relative importance for distinguishing between various types of beats. Hence we transform the components by subtracting the average and normalizing to a common standard deviation.

We have a mapping from a 12-dimensional input space to a 2-dimensional (5×5) output space. This is in contrast to many applications using SOMs, where the key goal is to reduce the dimensionality of the problem. In our case, that reduction has been achieved with the Hermite function decomposition. One central reason here for using SOMs is to benefit from its topological structure when interpreting the data.

4.2.2 Training

The training, which is divided into two phases, turns out to be robust with regard to all parameters within reasonable windows. In phase I the output vectors evolve with a strong neighbourhood dependence, whereas phase II proceeds with a weaker dependence. The latter allows for outliers to be identified, which are important for diagnosis. Below we describe in more detail the parameters and procedures we use.

Initializing the weights. The input data typically occupies a very small subspace of the total input space. Ideally, one wants the output nodes to focus on this subspace from the start. Since the input vector components have been normalized to zero mean, this is accomplished by initializing ω_{jk} to random[-ε,ε], where ε is chosen small compared to the standard deviation, $\varepsilon = 0.01$.

Local neighbourhood. In order to produce a good topological map, one should start with a wide neighbourhood (T_0 large), compared to the size of the network. This will force all units to become more or less similar. The neighbourhood is then decreased during training, allowing the units to focus on local regions of the input space. For a too small T_0, the coupling between the units is broken and no ordering will emerge. On the other hand, if T is large all the time, all units in the network will end up pointing in the same direction. We decrease the neighbourhood according to

$$T_{t+1} = cT_t \tag{16}$$

with $c = 0.95$. The initial T is estimated from the size of the output map as

$$T_0 = -\frac{d^2}{ln(2)} \tag{17}$$

where d denotes the diagonal distance of the map. This implies that a output vector at distance d from the winner will be updated with half the strength of the winner.

In phase I, T should not be too low. Hence T is chosen as

$$T = \max(T_{t-1}c, T_{\min}) \tag{18}$$

where $T_{\min} = 2/ln(2)$. In phase II there is no lower bound for T.

Learning rate. The learning rate η is in phase I annealed according to

$$\eta_{t+1} = k\eta_t \tag{19}$$

with $\eta_0 = 1$ and k = 0.96. In order for the convergence to be reasonable fast, η is chosen as

$$\eta_t = \max(\eta_{t-1}c, \eta_{\min}) \tag{20}$$

with $\eta_{\min} = 0.01$. In phase II η is fixed to $\eta_{\min}$.

The weights are updated for every pattern. Other updating rates have been pursued with less efficient training as results.

Termination criteria. The energy E of the system is defined as

$$E(t) = \sum_{ik} (\xi_k^i - \omega_{i^*k}(t))^2 \tag{21}$$

where t is the time step and i runs over all the input vectors. The fractional change in E from the last block of 10 time steps to the previous block defines ΔE. Both phases are terminated when $\Delta E < 0.001$.

5 Results

5.1 Resulting Maps

Depending upon the random number initialization, different features will after completed training settle in different parts of the 5×5 plane. However, the topological relations should be preserved. In Fig. 3 we show the beat assignments of the 5×5 map of output nodes, for one of the 30 minute records in the database. Each box represents a cluster of beats. Also indicated is the contents in terms of four kinds of beats; normal beats (N), atrial premature beat (A), ventricular premature beats (V) and fusion of ventricular and normal beats (F). In this case all 7 A and 11 F beats ends up in clusters where their kind are not dominant. Note that the beats plotted in the boxes are the original beats of lead B, whereas the input to the SOM are composed of the Hermite representation and the R-measures of both leads.

5.2 Evaluation of Results

The results above indeed look very good. We next turn to a more quantitative evaluation of the results. First we do this without benchmarking against other unsupervised approaches. Then we compare with the results of [15]. In Section 6 we will make comparisons with a cross-correlation clustering method.

A beat is defined as misclassified if it ends up in a cluster where the dominant beat is different. The resulting total percentage of misclassified beats for all ECG recordings is equal to 1.5%. The worst misclassification percentage is 12%, which would be obtained if all beats of each record were classified in a single box.

Results for 4, 5 and 6 Hermite functions decomposition are shown in Table 5. Table 6 shows the result divided into the different beat types. The majority of beats are normal (68.3%) and consequently the likelihood that normal beats will dominate the clusters is high. Most of the normal beats (99.8%) are assigned to clusters in which they represent the dominant fraction. The most common type of misclassification (45.7%) is that non-normal beats are assigned to clusters dominated by the normal ones. Another common type of misclassification is between classes that by definition are very close to each other. The f beats are a mixture between P (paced) and N (normal) beats and the F beats are a mixture between V (ventricular) and N beats. P beats classified as f or vice versa and V beats classified as F or vice versa account for 21.4% of all misclassifications. The unclassifiable beats (Q beats) constitute a very heterogeneous group and, not surprisingly, most of these beats are misclassified.

As mentioned in the introduction, in [15] the MIT-BIH Arrhythmia Data Base [16] was used for classifying beats using ANN methods. Not only do the overall objectives of [16] differ from ours.

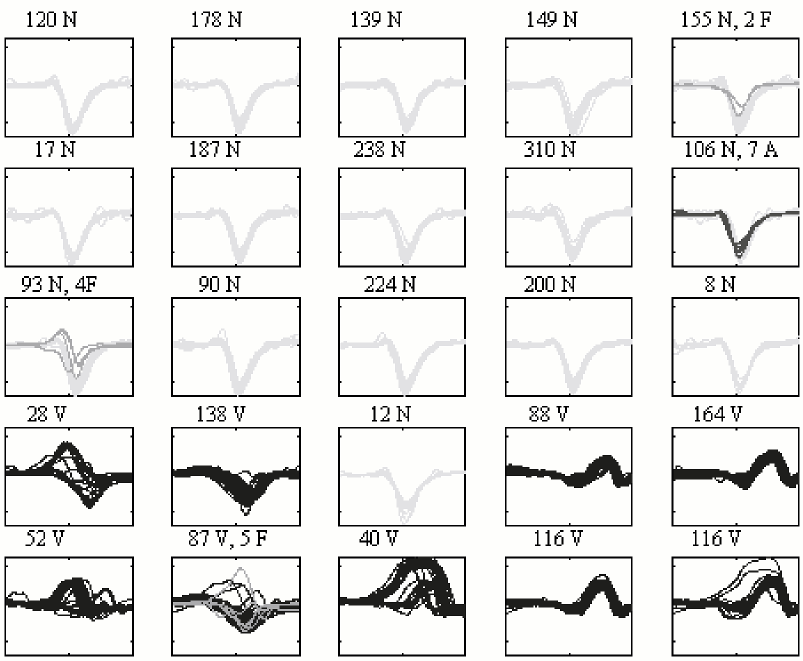

Figure 3: Beats from a typical 30 minute record (# 233 lead B) assigned to the different clusters (output nodes). For the notations N, A, V and F see Table 1.

	4 Hermite fcn	5 Hermite fcn	6 Hermite fcn
SOM	1.55 %	1.51 %	1.54 %

Table 5: Misclassifications in percentage averaged over all the available beats for different number of Hermite functions.

They also use a different methodology, which does not identify the QRS complexes (the database annotations are used) and, in addition, initial labeling of the beats is required. This is contrast to our approach, which handles all the three steps, QRS extraction, compact representation and classification, in an integrated manner. With regard to the methods used, the representation of each beat is different. Whereas [15] uses the 9 largest principal components computed from 14 samples (180Hz) on each side of the R-peak, we decompose the beats into Hermite functions. Another difference is that we use all 48 records in contrast to [15], where only 33 records (20 for the evaluation), in which premature ventricular contraction can be identified, are kept. Also, the authors of [15] limit their study to a two compressed classes, V and non-V respectively. For the classification, the authors of [15] use a MOE variant of learning vector quantization, whereas in this work we are consistently employing unsupervised learning (clustering). We next compare our results for the

	N	L	R	A	a	J	S	V	F	b	e	j	E	P	f	Q	$\tilde{\Sigma}$
N	74699	0	2	49	1	12	0	49	8	0	0	31	0	0	3	0	155
L	0	8046	0	6	0	0	0	4	0	4	0	0	0	0	0	0	14
R	1	0	7095	139	0	0	0	0	0	4	0	0	0	0	0	0	144
A	281	6	34	2213	0	0	0	4	0	0	0	0	0	0	0	0	325
a	13	0	0	9	57	0	0	30	0	0	0	0	0	0	0	0	52
J	7	0	29	0	0	47	0	0	0	0	0	0	0	0	0	0	36
S	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
V	186	3	0	5	1	0	0	6811	79	23	0	0	1	1	0	0	299
F	146	1	3	0	0	0	0	49	599	0	0	0	0	0	0	0	199
b	0	0	0	4	0	0	0	7	0	446	0	0	1	0	0	0	12
e	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
j	75	0	6	0	0	0	0	0	0	0	0	147	0	0	0	0	81
E	1	1	22	0	0	0	0	0	0	16	0	0	66	0	0	0	40
P	1	0	0	0	0	0	0	3	0	0	0	0	0	6942	68	0	72
f	15	0	0	0	0	0	0	1	0	0	0	0	0	158	805	0	174
Q	11	2	0	0	0	0	0	2	0	0	0	0	0	5	11	2	31
$\tilde{\Sigma}$	755	13	96	212	2	12	0	149	87	47	0	31	2	164	82	0	

Table 6: The SOM clustering of the different types of beats using 5 Hermite functions for the representation and a 5×5 output map. The result from the individual files are added together. The annotation file types are depicted on the vertical axis whereas the clusters dominated by the different types are shown horizontally. For example, 31 of the N beats are assigned to clusters where j is dominant. $\tilde{\Sigma}$ denotes summed rows and columns excluding the diagonal.

classification part with those of [15].

We process the same records as in [15] on a 5×5 map. In order to make comparisons we need to define a proper error measure. In [15] a performance measure is defined as

$$R^{\mathrm{ME}} = \frac{\#\text{ of true negative and true positive beats}}{\text{total }\#\text{ of beats}} \tag{22}$$

For comparison we define a similar measure, noting that V in [15] correspond to our V and E,

$$R^{\mathrm{SOM}} = \frac{\text{number of correctly clustered V and E beats}}{\text{total }\#\text{ of beats}} \tag{23}$$

The comparison between our SOM and the results from MOE [15] are found in Table 7. If an expert were to annotate one typical beat for each of the 25 clusters in our output map while agreeing with the data base annotation, one could consider the beats in that cluster as classified to that type. As can be seen from Table 7, within this scenario our method outperforms their method. As a reference a column with the result obtained if all beats where classified as normal (worst case for the SOM approach) is included.

record	R^{SOM} (%)	R^{ME}(%)	ref. (%)
200	98.1	81.0	68.0
201	99.9	95.4	89.7
202	99.4	71.7	99.1
203	96.8	87.4	85.1
205	99.9	97.1	97.4
207	96.9	88.3	90.9
208	98.5	90.5	66.3
209	100.0	98.7	100.0
210	98.3	93.4	92.7
213	97.6	91.9	93.2
214	99.9	98.2	88.7
215	99.9	98.4	95.1
219	99.6	97.4	97.0
221	100.0	99.0	83.7
223	99.1	94.4	81.9
228	100.0	99.0	82.3
230	100.0	99.1	100.0
231	99.9	99.8	99.9
233	99.8	97.7	73.0
234	100.0	99.8	99.9
Avg.	99.1	94.0	88.8

Table 7: Comparison of the mixture of expert model [15] using the R-measures defined in (22) and (23). The reference column "ref." shows the extreme case of classifying all beats as normal.

6 Crosscorrelation-Based Method

We also gauge our self-organizing approach against a conventional method based on crosscorrelation (**CC**) coefficient and a set of template beats. This method can, similar to the SOM, be categorized as unsupervised classification. Due to memory constraints when implementing the CC method, each of the 48 records in the database is divided into three 10 minutes long records.

The CC method is initiated by using the first beat as a template beat. Each beat is subjected to linear-phase, bandpass filtering in order to remove frequencies which are judged to be less important for the classification procedure (filter cut-off frequencies at 1 and 35 Hz). The subsequent beats are then compared to the template beat with respect to the CC coefficient; each beat is shifted in time until the highest CC coefficient is found. A new template beat is created whenever the CC coefficient drops below a noise-dependent threshold value. The threshold is set to 0.94 at low to moderate noise levels while then gradually decreasing to 0.75 at high noise levels; this threshold design ensures the creation of new beat classes in noisy signals is restricted. The noise level is estimated from the RR interval prior to the QRS complex. A beat classified as being similar to an existing class is then used to update the template beat by means of recursive averaging. This procedure will successively improve the quality of the template beats.

For each 10 minutes record we train a SOM with the same number of clusters ($A \times B$) as produced by the CC method. We compute the total percentage of misclassified beats for both methods (as

Method	# misclassified
Correlation	4.4%
SOM $A \times B$	2.3%
SOM 5×5	1.3%

Table 8: Results of the 10 minutes recording used to compare SOM with the correlation method. The SOM is evaluated for a 5x5 map and for the size produced by the correlation method, $A \times B$.

defined above). The same is done with fixed map sizes of 5×5. The results from the CC method is 4.4% misclassified, which should be compared with 2.3% for the SOM under similar conditions.

7 Summary

We have devised a procedure for clustering beats into classes which are not predefined. The method entails Hermite function representation and self-organized neural networks for the purpose of beat clustering. It is found that the QRS complex of the beats in the MIT-BIH Arrhythmia Database are efficiently extracted (99.7%). Decomposing the beats into five Hermite functions turns out to be sufficient for achieving a good classification performance. The entire approach is successfully evaluated in three different and independent ways using 5×5 output maps:

- The degree of misclassification is very low (1.5%), where the former is defined by beats ending up in clusters where the dominant kind is another one. Furthermore, when analyzing the resulting topological maps, these tend to cluster the beats such that similar features are adjacent. The CPU consumption needed for the clustering is modest, less than 1 minute per record on an desktop computer (Digital Alpha 250).
- When comparing the clustering results with those from supervised learning methods [15] by classifying each of the 25 clusters according the dominant kind of beat, our method outperforms the supervised approach.
- As a reference we also classify 10 minutes long records using a method based on template cross-correlation [11]. It turns out that our method performs significantly better than the latter.

One should mention that our approach not only classifies well; it also contains topological information, which is still to be completely exploited. Needless to say, it is very convenient for the cardiologist if similar clusters are presented as neighbours in the map.

It should be noted that our approach embraces all processing steps from the raw ECG signal to the clustered beats; no corrections based on annotations were inserted in between the different steps.

The performance based on an input vector to the SOM which does not include any measurements on signal quality, e.g. information on signal-to-noise ratio. It is reasonable to assume that the performance will become even better when such measurements are included. Another step which remains to be considered is the classification of beat episodes, which would rely on the present classification of single beats combined with information on rhythm.

It is not within the scope of this project to adjust the method to perform particularly well for a certain type of beats. The best performance is found for the normal beats, a class that also outnumbered the other beats. However, if it for medical reasons is considered more important to find for example the ventricular beats it is possible to change the importance of width measures by e.g. simply multiplying that input feature by a factor larger than one.

Acknowledgements: This work was supported by the Swedish National Board for Technical Development and by the Swedish Medical Research Council. We also acknowledge Ola Rauer and Ralf Rittner for contributions at an early stage of this project.

References

[1] J.L. Willems, C. Abreu-Lima, P. Arnaud, L.H. van Bemmel, C. Brohet, R. Degani, B. Denis, J. Gehring, I. Graham, G. van Herpen, H. Machado, P.W. Macfarlane, J. Michaelis, S.D. Moulopoulos, P. Rubel, C. Zywietz, "The Diagnostic performance of computer programs for the interpretation of electrocardiograms", *New England Journal of Medicine* **325**, 1767-1773 (1991).

[2] W.G. Baxt, "Application of artificial neural networks to clinical medicine", *Lancet* **346**, 1135-1138 (1995).

[3] B. Hedén, H. Öhlin, R. Rittner, L. Edenbrandt, "Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks", *Circulation* **96**, 1798-1802 (1997).

[4] H.-S. Chow, G.B. Moody, R.G. Mark, "Detection of ventricular ectopic beats using neural networks", *Proc. of Computers in Cardiology, IEEE Computer Society*, 659-662 (1992).

[5] M. Morabito, A. Macerata, A. Taddei, C. Marchesi, "QRS morphological classification using artificial neural networks", *Proc. of Computers in Cardiology, IEEE Computer Society*, 181-184 (1991).

[6] R. Watrous, G. Towell, "A patient-adaptive neural network patient monitoring algorithm", *Proc. of Computers in Cardiology, IEEE Computer Society*, 229-232 (1995).

[7] P. Bozzola, G. Bortolan, C. Combi, F. Pinciroli, C. Brohet, "A hybrid neuro-fuzzy system for ECG classification of myocardial infarction", *Proc. of Computers in Cardiology, IEEE Computer Society*, 241-244 (1996).

[8] T.Y. Young and W.H. Huggins, "On the representation of electrocardiograms" *IEEE Trans. Biomed. Eng.* **BME-10**, 86-95 (1963).

[9] S. Karlsson, "Representation of ECG records by Karhunen-Loeve expansion", *Dig. 7th Int. Conf. Med. & Biol. Eng., Stockholm*, 105 (1967).

[10] N. Ahmed, P.J. Milne and S.G. Harris, "Electrocardiographic data compression via orthogonal transforms", *IEEE Trans. Biomed. Eng.* **BME-22**, 484-487 (1975).

[11] L. Sörnmo, P.O. Börjesson, M.E. Nygårds and O. Pahlm, "A method for evaluation of QRS Shape features Using a Mathematical Model for the ECG", *IEEE Trans. Biomed. Eng.* **BME-28**, 713-717 (1981).

[12] P. Laguna, R. Jane, P. Caminal, H. Rix and N.V. Thakor, "Adaptive estimation of the QRS complex wave in the electrocardiographic signal (ECG) by the Hermite model: Classification and Ectopic Beat Detection", *Med. Biol. Eng. Comput.* **34**, 58-68 (1995).

[13] R. Jane, S. Olmos, P. Laguna and P. Caminal, "Adaptive Hermite models for ECG data compression: performance and evaluation with automatic wave detection", *Proc. of Computers in Cardiology, IEEE Computer Society*, 389-392 (1993).

[14] T. Kohonen, "Self organized formation of topologically correct feature maps", *Biolological Cybernetics* **43**, 59 (1982); T. Kohonen, *Self-organization and Associative Memory*, third edition, Springer-Verlag, Berlin, Heidelberg (1990).

[15] Y.H. Hu, S. Palreddy and W. J. Tompkins, "A Patient-adaptable ECG beat classifier using a mixture of experts approach", *IEEE Transactions on Biomed. Eng.* **44**, 891-900 (1997).

[16] R. Mark and G. Moody, "MIT-BIH arrhythmia data base directory", Massachusetts Institute of Technology (1988).

[17] M.E. Nygårds, L. Sörnmo, "Delineation of the QRS complex using the envelope of the ECG" *Medical and Biological Engineering & Computing*, **21**, 538-547 (1983).